

# Wikidata and Archival Metadata

## opportunities and challenges for heritage institutions

Study day on Linked Data (#LinkingThePast)

[“Linking the Past : Possibilities and limits of publishing archival data as Linked Data”](#), November 22, 2019. KBR, Brussels

Baptiste de Coulon (le lieu imaginaire) [@b3d2c](#) / [@b3d2c@framapiaf.org](mailto:@b3d2c@framapiaf.org)

This presentation and more on : <https://frama.link/LkPast>

# Acknowledgment

Many thanks to : Florence Gillet (State Archives in Belgium) and  
Anne Chardonnes (Free University of Brussels)  
for their invitation.

It's a great pleasure to be here !

A grateful thank-you to my sister Pamina de Coulon who translated a large part of this presentation into English. The remaining faults are my responsibility.

And, finally, thanks to [Harmonia Amanda](#) who hosted the first Wikidata workshop in Romandie (the French-speaking part of Switzerland) in 2018 and, on that occasion, shared with us some of her superpowers.

0. **Who** I am?

1. **What** is Wikidata?

2. **Why** add information about our archives to Wikidata?

3. **How** to add information about our archives to Wikidata?

## 0. Who I am?

I am a swiss archivist and Wikimedian.

I run my own business call “le lieu imaginaire<sup>1</sup>” but I am also employed partial time at the Manuscript’s Department of the Library of the University of Lausanne<sup>2</sup>.

1 <https://lelieuimaginaire.ch/>

2 <https://www.bcu-lausanne.ch/patrimoine/collections-precieuses/manuscrits/#.WtdFHNYuBhE>

# 1. What is Wikidata?

Wikidata is the first project on the web more powerful than the research engine of Google (1998). We waited 20 years for this. You can find on it preciser answer on general theme than anywhere else.

Wikidata<sup>3</sup> is a project of the Wikimedia Foundation lead by Wikimedia Deutschland. It was launched in 2013.

It's a knowledge base build on a centralized repository  
collaboratively edited online.

The use is free.

The data are structured according to Resource Description Framework (RDF) specifications and published under the CC0 free license.

Wikidata do not store original information.

→ All statements (entity–attribute–value) must have references.

<sup>3</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

# 1. What is Wikidata?

From our point of view, Wikidata will be the next key tool for heritage institution's professionals, if it is not already.

We see three main uses :

1. find information (Do you speak SPARQL<sup>4</sup>?)
2. clean and link authorities (Do you know Openrefine<sup>5</sup>?)
3. give visibility to their collections (Why not?)

4 [https://www.wikidata.org/wiki/Wikidata\\_talk:Events/AAFSaint%C3%8919#SPARQL](https://www.wikidata.org/wiki/Wikidata_talk:Events/AAFSaint%C3%8919#SPARQL) (introduction)

5 <http://openrefine.org/>

## 2. Why add information about our archives to Wikidata?

This being the main question raised by our colleagues, our friends, and maybe even by you, we will start from it.

To begin answering this question, we must ask ourselves the question of resources:

to work with Wikidata means not doing something else.

It cannot just be added to the things we are already doing, it has to replace some that we do.

## 2. Why add information about our archives to Wikidata?

Hence what we propose as answers :

1. because we want to reach a larger audience, and therefore exit our institutional boundaries and national or language border.
2. because we don't want to keep on writing up information that already exists.
3. because we want our data to be easily to reuse.
4. At last, because it's important but also fun, challenging, powerful and so useful.



## 2. Why add information about our archives to Wikidata?

1. because we want to reach a larger audience, and therefore exit our institutional boundaries and national or language border.

Well, the public is on the Wikipedia(s).

No doubt : we are all also!

More than 300 different languages versions of Wikipedia, which regularly make it the 5th most visited website in the world.

And since Wikidata is becoming the central piece of the Wikimedia projects, it will become the driving force of future infoboxes of Wikipedias.

## 2. Why add information about our archives to Wikidata?

Example : Marie Heim-Vögtlin ([Q117598](#)) / Statement “archived by” (P485)



The screenshot shows the Wikidata interface for the property 'archives at' (P485) for the entity Marie Heim-Vögtlin (Q117598). The interface is divided into two main sections, each representing a different archive. Each section includes the name of the archive, an 'inventory number' field, and a '1 reference' link. An 'edit' button is visible next to the archive name in each section. At the bottom right of the interface, there is a '+ add value' button.

Archive Name	Inventory Number	Reference Count
ETH Zurich University Archives ...	CH-001807-7:Hs 724 ...	1 reference
Archive for the History of Medicine ...	PN 198 ...	1 reference

## 2. Why add information about our archives to Wikidata?

[Wikipedia in Catalan](#) {{Infotaula persona}} | [Wikipedia in French](#) {{Infobox Biographie2}}

**Marie Heim-Vögtlin**

**Biografia**

**Naixement** (de) Marie Vögtlin  
7 octubre 1845  
Bözen (Suïssa)

**Mort** 7 novembre 1916 (71 anys)  
Zuric (Suïssa)

**Lloc d'enterrament** Cementiri Sihfeld (Zuric)

**Formació** Universitat de Zuric

**Activitat**

**Ocupació** Metgessa i ginecòloga

**Obra**

**Localització dels arxius**

- ETH Zurich University Archives (CH-001807-7:Hs 724)
- Archive for the History of Medicine (PN 198)

**Família**

**Cònjuge** Albert Heim

**Fills** Arnold Heim

**Parents** Friedrich Erismann (cosí germà)

**Marie Heim-Vögtlin**

**Biographie**

**Naissance** 7 octobre 1845   
Bözen

**Décès** 7 novembre 1916  (à 71 ans)  
Zurich

**Sépulture** Friedhof Sihlfeld (d)

**Nom de naissance** Marie Vögtlin

**Nationalité** Suisse

**Formation** Université de Zurich

**Activités** Médecin, gynécologue

**Conjoint** Albert Heim

**Enfant** Arnold Heim

**Parentèle** Friedrich Erismann (en) (cousin germain)

**Autres informations**

**Archives conservées par** Archives de l'École polytechnique fédérale de Zurich (en) (CH-001807-7:Hs 724)  
Archiv für Medizingeschichte, Lehrstuhl Medizingeschichte (d) (PN 198)

modifier - modifier le code - modifier Wikidata

## 2. Why add information about our archives to Wikidata?

2. because we don't want to keep on writing up information that already exists.

Wikidata cover :

1. about 70 million item's Wikipages (see statistics<sup>6</sup>)

for about 2 million pages in the french version of Wikipedia

2. soon 5.8 million human beings

of which less than 20% are of feminine gender !

for about 600'000 in the french version of Wikipedia

and 3.7 million in all the Wikimedia projects

Therefore:

- › either the information already exists
- › or we have the privilege of creating it

<sup>6</sup> <https://www.wikidata.org/wiki/Wikidata:Statistics/en>

## 2. Why add information about our archives to Wikidata?

3. Because we want our data to be easy to reuse.

From Wikidata,

- the data are openly downloadable
- under license CC0 (like public domain)
- in structured formats
- mass exports are possible
- an API (application programming interface) allows interconnections (for instance with Openrefine)

## 2. Why add information about our archives to Wikidata?

4. because it's important but also fun, challenging, powerful and so useful.

- Important because worldwide GLAM have the new mission to build a digital free knowledge infrastructure and Wikimedia Foundation is our main partner for that<sup>7</sup>.
- Fun, challenging, powerful because Wikidata's tools are update and user-friendly. Developer's team is listening to users<sup>8</sup>.
- So useful : feedback from the public are so grateful!

7 <https://en.wikipedia.org/wiki/Wikipedia:GLAM>  
[https://meta.wikimedia.org/wiki/Strategy/Wikimedia\\_movement/2018-20](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20)

8 <https://phabricator.wikimedia.org>

### 3. How to add information about our archives to Wikidata?

We have participated to establish a documentation concerning archives (see [WikiProject Archival Description](https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description)<sup>9</sup>).

It is still being developed, so it's incomplete, messy and yet to be translated into English, but it exists.

Where to start ?

1. Describe your institution in a Wikipage of Wikidata if not already the case.
2. Focus your work on authority record.
3. Once this is done, you will still be able to put the information about your collections themselves but we will maybe have the opportunity to talk about it again until then :-)

<sup>9</sup> [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Archival\\_Description/Data\\_structure](https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description/Data_structure)

### 3. How to add information about our archives to Wikidata?

1. Describe your institution in a Wikipage of Wikidata (with a Qxxxx ID)  
See the documentation<sup>10</sup>.

2. Then, we propose you focus your work on authority record.

*Among these, you can privilege those of feminine gender, which will contribute to reduce the gender gap.*

In order to do that, you have :

- first to establish the list of your authority record
- verify their existence in Wikidata (via Openrefine for example)
- add the references of the archives that you have about them thanks to the “archived by” property ([P458](#)<sup>11</sup>).

<sup>10</sup> [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Archival\\_Description/Data\\_structure#Service\\_administratif](https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description/Data_structure#Service_administratif)

<sup>11</sup> [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Archival\\_Description/Data\\_structure#P485](https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description/Data_structure#P485) (archivé par)



3. Once this is done, you will still be able to put the information about your collections themselves

*Example : the pilot project<sup>12</sup> lead by Dominic Byrd-McDevitt for National Archives and Records Administration (NARA):*

*43 organizations*

*571 Records Groups*

*4'754 Collections*

*429'295 items identified in the NARA online catalog, including:*

*284'982 Collections (Q2668072)*

*118'323 item of collection or exhibition (Q18593264)*

<sup>12</sup> [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Archival\\_Description/Use\\_cases#NARA](https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description/Use_cases#NARA)

So, when do we start?

# Making-off

Writing using LibreOffice (Ubuntu 18.04) with Fira Sans typeface<sup>13</sup>

most of the time in the train.

<sup>13</sup> [https://en.wikipedia.org/wiki/Fira\\_Sans](https://en.wikipedia.org/wiki/Fira_Sans)